International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR)
ISSN(P): 2249-6831; ISSN(E): 2249-7943
Vol. 4, Issue 1, Feb 2014, 125-132

IMPROVE THE TECHNIQUE OF RELEVANCE FEEDBACK FOR CONTENT-BASED MULTIMEDIA ARCHIVING BY USING APRIORI ALGORITHM

AASMA MUJAWAR & S .P.KOSBATWAR

Smt. Kashibai Navale College of Engineering, Vadgaon (Bk), Pune, Maharashtra, India

ABSTRACT

To provide efficient and effective retrieval of content based multimedia data and images from multimedia database like video data, images by using relevance feedback technique and mining algorithm. By using NPRF, high quality of image retrieval on RF can be achieved in a small number of feedback. Proposed algorithm NPRF Search performs the navigation-pattern-based search to match the user's intention by merging three query refinement strategies. As a result, traditional problems such as visual diversity and exploration convergence can solve. Efficiency of content-based multimedia retrieval can measure in terms of following factors precision, converge and number of feedbacks.

KEYWORDS: Content, Based Multimedia Retrieval, Relevance Feedback, Query Image Reweighting, Query Expansion, Query Point Movement

INTRODUCTION

In many areas of commerce, government, academia, and hospitals, large collections of digital images are being created. Many of these collections are the product of digitizing existing collections of analogue photographs, diagrams, drawings, paintings, and prints. Usually, the only way of searching these collections was by keyword indexing, or simply by browsing. Digital images databases however, open the way to content-based searching. Mining Images is the knowledge discovery from the Image Database. Image Retrieval is an active area of research with many applications such as Image Browsing. Content Based Image Retrieval has been an interesting subject of many researchers in recent years. There are many great developing Retrieval approaches and techniques to improve the Retrieval accuracy. Project Approach is all about text retrieval from the combinatorial textual image database. Itis very important to mine the images in order to find hidden information within the Images. As well as search those Images to provide required one to user in lesser time. This Intended method is basically conceded into two Parts: Text Extraction and Textual Image Retrieval using Sketch. These Two parts are implemented with help of "Algorithm for Text Extraction", and "Algorithm for Textual Image Retrieval Using Sketch" respectively. Proposed Method present effective as well as robust text extraction algorithms, since it automatically detect and extract text from scene, caption, and different textual images. The contributions of the projected method are, they can handle both printed document and scene text images, with not more sensitive to image color/intensity, robust with respect to font, sizes, orientations, and distinguishes text regions from texture-like regions, suchas window frames, wall patterns, etc. Also it provides fast retrieval of Images to the Query by Sketch from the user. This is a very efficient way of searching. The purpose of this document is to enhance retrieval technique of images and videos by reducing number of iterations. [8]In a traditional database system, the result of a query is a set of values (those values that satisfy the query). In other data servers, such as a system with queries based on image content, or many text retrieval systems, the result of a query is a sorted list. For example, in the case of a system with queries based on image content, the query might ask for objects that are a particular shade of red, and the result of the query would be a sorted list of objects in

the database, sorted by how well the color of the object matches that given in the query. A multimedia system must somehow synthesize both types of queries (those whose result is a set and those whose result is a sorted list) in a consistent manner. In this paper we discuss the solution adopted by Garlic, a multimedia information system being developed at the IBM Almaden Research Center. This solution is based on “graded” (or “fuzzy”) sets. Issues of efficient query evaluation in a multimedia system are very different from those in a traditional database system. This is because the multimedia system receives answers to subqueries from various subsystems, which can be accessed only in limited ways. For the important class of queries that are conjunctions of atomic queries (where each atomic query might be evaluated by a different subsystem), the naive algorithm must retrieve a number of elements that is linear in the database size. In contrast, in this paper an algorithm is given, which has been implemented in Garlic, such that if the conjuncts are independent, then with arbitrarily high probability, the total number of elements retrieved in evaluating the query is sublinear in the database size (in the case of two conjuncts, it is of the order of the square root of the database size). It is also shown that for such queries, the algorithm is optimal. The matching upper and lower bounds are robust, in the sense that they hold under almost any reasonable rule (including the standard min rule of fuzzy logic) for evaluating the conjunction. Finally, we find a query that is provably hard, in the sense that the naive linear algorithm is essentially optimal.[4] Relevance feedback has emerged as a powerful tool to boost the retrieval performance in content-based image retrieval (CBIR). In the past, most research efforts in this field have focused on designing effective algorithms for traditional relevance feedback. Given that a CBIR system can collect and store users' relevance feedback information in a history log, an image retrieval system should be able to take advantage of the log data of users' feedback to enhance its retrieval performance. In this paper, we propose a unified framework for log-based relevance feedback that integrates the log of feedback data into the traditional relevance feedback schemes to learn effectively the correlation between low-level image features and high-level concepts. Given the error-prone nature of log data, we present a novel learning technique, named Soft Label Support Vector Machine, to tackle the noisy data problem. Extensive experiments are designed and conducted to evaluate the proposed algorithms based on the COREL image data set. The promising experimental results validate the effectiveness of our log-based relevance feedback scheme empirically.

EVALUATION

A **Relevance-Feedback** based approach to CBIR, in which a computer interacts to refine high-level queries to representations based on low-level features. Relevance feedback is a powerful technique used in traditional text-based information retrieval systems. It is the process of automatically adjusting an existing query using the information fed back by the user about the relevance of previously retrieved objects such that the adjusted query is a better approximation to the user’s information need. In the relevance-feedback-based approach, the retrieval process is interactive between the computer and the human. Under the assumption that high-level concepts can be captured by low-level features, the relevance feedback technique tries to establish the link between high level concepts and low-level features from the user’s feedback. Furthermore, the burden of specifying the weights is removed from the user. The user only needs to mark which images he or she thinks are relevant to the query. It has proven to be very effective for improving retrieval performance. Relevance feedback is used to reduce semantic (meaning) gap in images.

Low level features:

- Color
- Texture
- Shape

High level concepts: User intention or requirement. (Example: **Fresh** apple)

Relevance feedback is used to reduce semantic (meaning) gap in images. That is mapping of low level features and high level concepts.

Query- Image Reweighting

In previous work contain what image features are important for those images (relevant images) select up by the users at each feedback (iteration). The notion behind QR is that, if the ith feature fi exists in positive (relevant) examples frequently, the system assigns the higher degree to fi. QR like approaches were first proposed by Rui et al. [2], which convert image feature vectors to weighted-term vectors in early version of Multimedia Analysis and Retrieval System (MARS). In query reweighting, feature weights are keep on changing to connect with high level user concepts and low level image features.

Query -Image Point Movement

To improve the accuracy of image retrieval is moving the query point towards the point of the user's intension in feature space. It gives positive(relevant) examples as a new query point at each iterations. After changing or updating new query point at each feedback , the query point should be move to a region of the user's requirement. The well-known space-vector formula proposed by Rocchio[2] is as follows:

$$Q_i = Q_{i-1} + \alpha \sum_{j=1}^{nr} R_j / nr - \beta \sum_{j=1}^{nir} IR_j / nir$$

where,

Qi is the vector of the i th query.

IRj is the vector of the jth irrelevant image.

nr is cardinality of relevant images.

nir is cardinality of irrelevant images.

Query-Image Expansion (QEX)

Query Reweighting and Query Point Movement cannot increase the quality of Relevance Feedback; Query Expansion has been another technique in the solution to increase quality of image retrieval recently. The search strategies, such as QR and QPM, cannot completely satisfy the user's interest to extract relevant images .Using Query Expansion, user picked image feature added into image resultset at each level of iteration . For this reason, the modified version of MARS [9] groups the similar relevant points into several clusters, and selects good points from these cluster to the multipoint query.

Initial Query Processing Phase

Without considering the feature weight, this phase extracts the visual features from the original query image to find the similar images. Afterward, the good examples (also called positive examples) picked up by the user are further analyzed at the first feedback (also called iteration 0).

Image Search Phase

Behind the search phase, our intent is to extend the one search point to multiple search points by integrating the navigation patterns and the proposed search algorithm NPRF Search. Thus, the diverse inclusion of the user's interest can be successfully implied. In this phase, a new query point at each feedback is generated by the preceding positive examples. Then, the k-nearest images to the new query point can be found by expanding the weighted query. The search procedure does not stop unless the user is satisfied with the retrieval results

Knowledge Discovery Phase

Learning from users behaviours in image retrieval can be viewed as one type of knowledge discovery. Consequently, this phase primarily concerns the construction of the navigation model by discovering the implicit navigation patterns from users' browsing behaviours. This navigation model can provide image search with a good support to predict optimal image browsing paths.

Data Storage Phase

The databases in this phase can be regarded as the knowledge marts of a knowledge warehouse, which store integrated, time-variant, and non-volatile collection of useful data including images, navigation patterns, log files, and image features. The knowledge warehouse is very helpful to improve the quality of image retrieval. Note that the procedure of constructing rule base from the image databases can be conducted periodically to maintain the validity of the proposed approach. for completing the specific process. For online operation, once a query image is submitted to this system, the system first finds the most similar images without considering any search strategy, and then returns a set of the most similar images.

The first query process is called **initial feedback**. Next, the good examples picked up by the user deliver the valuable information to the image search phase, including new feature weights, new query point, and the user's intention. Then, by using the navigation patterns, three search strategies, with respect to QPM, QR, and QEX, are hybridized to find the desired images. Overall, at each feedback, the results are presented to the user and the related browsing information is stored in the log database. After accumulating long-term users' browsing behaviours, offline operation for knowledge discovery is triggered to perform navigation pattern mining and pattern indexing.

SYSTEM ARCHITECTURE

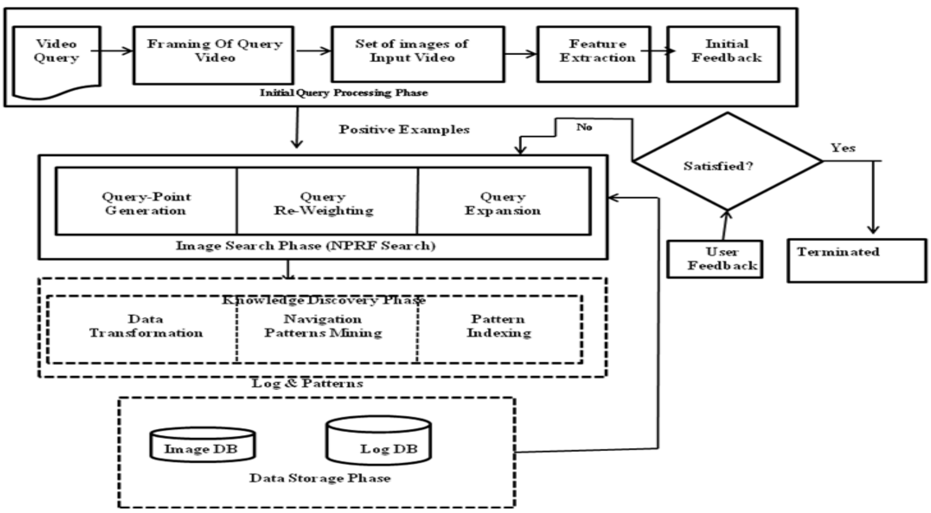

Figure 1

The major difference between our proposed approach and other approaches is that we approximate an optimal solution to resolve the problems existing in current RF, such as **redundant browsing and exploration convergence**. To this end, the approximated solution takes advantage of exploited knowledge (navigation patterns) to assist the proposed search strategy in efficiently hunting the desired images. Generally, the task of the proposed approach can be divided into two major operations, namely offline knowledge discovery and online image retrieval. Each operational phase contains some critical components. This architecture focuses on the discovery of relations among the users' browsing behaviors on RF. Basically; the frequent patterns mined from the user logs are regarded as the useful browsing paths to optimize the search direction on RF. In our NPRF approach, the users' common interests can be represented by the discovered frequent patterns (also called frequent itemsets). Through these navigation patterns, the user's intention can be precisely captured in a shorter query process. In this phase, the Apriori-like algorithm is performed to exploit navigation patterns using the transformed data. The task for establishing the navigation model can be decomposed into two steps:

Step 1: Construction of the navigation transaction table. From Fig. 3, let us select five query sessions as an example shown in Table 1. In Table 1, a query session can be considered a transaction. In this case, the transaction is composed of a query item and several iteration items To exploit valuable navigation patterns, all query sessions in the transformed log table are collected as the **navigation-transaction table.** From Table 1, we can discover some interesting phenomena. First, though passing through different iterations, the paths starting with the same query item lead to the same destination, e.g., Session 001 and Session 002. Second ,the paths starting with different query items lead to the same destination, e.g., Session 002 and Session 004. Third, the paths starting with the same query item lead to different destinations, e.g., Session 003 and Session 004. Even the special path, Session 005, may be the other important trail for image hunting. Such evidence implicates the main aspect we want to demonstrate in this example.

Step 2: Generation of navigation patterns. This operation concentrates on mining valuable navigation patterns to facilitate online image retrieval. As shown in Table 1, the frequent itemsets X whose supports support(X) exceed the presetting minimum support minsup are mined by Apriori-like algorithm. In fact, the generated frequent itemsets can be regarded as sequential navigation patterns directly since temporal continuities have been considered in creating QPD. For example, as shown in Table 1, the sequential navigation pattern f C11 C32 C42 derived from frequent itemset fC11;C32;C42 under the minimum support is 2.

Table 1: Example of Navigation Patterns

Query Session ID	Item
001	C_{11},C_{21},C_{32},C_{42}
002	C_{11},C_{23},C_{32},C_{42}
003	C_{12},C_{21},C_{31},C_{42}
004	C_{12},C_{21},C_{31},C_{42}
005	C_{13},C_{22},C_{32},C_{43}

Navigation Pattern Search Algorithm Steps

Step 1. Initially feature extraction of all the images of the database is performed.

Step 2. User inputs a query image.

Step 3. Feature extraction of query image is performed.

Step 4. Relevance feedback technique is implemented using query reweighting, query expansion, query point movement done.

Step 5. Result is displayed.

Step 6. If user satisfied then searching get finished.

Step 7. Else repeat process of query reweighting, query expansion, query point movement technique.

Step 8. User puts his feedback after examining the searched result.

Step 9. User repeats this step until and unless he is not satisfied.

SYSTEM OVERVIEW

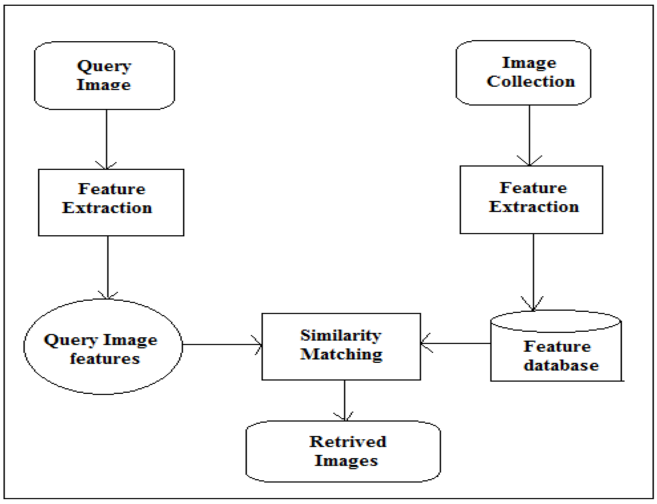

Figure 2: System Flow Diagram

Figure shows the System flow diagram. Here system flow diagram for Image Retrieval process shown. First we collect set of images as Dataset. Features extraction of images will be implemented using color and texture formula. Matrix formation of content-based images. Then grouping of set of relevant as well as irrelevant images done by k-means algorithm. This will off line process. When user fired a query as image. It uses these cluster of images to retrieve most relevant of set of images. In fist iteration we will get set of relevant and irrelevant images. Then by using these techniques query point movement ,query generation we will get user satisfied set of images. This process will done until we will get most relevant set of images. This process will done by apriori algorithm to reduced number of iterations.

CONCLUSIONS

The proposed navigation pattern relevance feedback technique of image and video retrieval improves retrieval quality. At each iteration, maintain list of user selected images to achieve valuable set of images and videos. It optimizes the retrieval quality of interactive CBIR..Reduces the iteration while extracting images from given dataset. Improves its performance .Maintain logs of images to satisfy user intention.

REFERENCES

1. Ja-Hwung Su, Philip S. Yu, Wei-Jyun Huang, ”Efficient Relevance Feedback for CBIR by Mining User Navigation Patterns” IEEE Trans On Knowledge And Data Engineering VOL. 23,No March 2011.
2. Mianchu Chen, Ping Fu, Yuan sun, Hui zhang ”Image Retrieval Based on Multi-feature similarity score fusion

using Genetic Algorithm”_The 2nd International Conference on Computer and Automation Engineering (ICCAE), 2010.

3. Yu Xiaohong, Xu Jinhua, "The Related Techniques of Content-based Image Retrieval", 2008 International Symposium on Computer Science and Computational Technology.
4. Ryszard S. Choras, "Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems", International, Journal of biology and biomedical engineering, Issue 1, Vol. 1, 2007
5. Alaa H. Halawani, Alexandra Teynor, Lokesh Setia, Gerd Brunner, Hans Burkhardt, "Fundamentals and Applications of Image Retrieval", An Overview Datenbank-Spektrum 18/2006.
6. Tobias Weyand, Thomas Deselaers," Combining Content-based Image Retrieval with Textual Information Retrieval", Department of Computer Science, RWTH, Aachen, 2005.
7. Fahui Long, Hongjiang Zhang, David Dagan Feng, "Fundamentals of content-based image retrieval", Microsoft corporation research articles, 2003. [6] Roger Weber, Michael Mlivoncic, "Efficient Region-Based Image Retrieval", ACM CIKM ‘03, pp. 3-8, USA, 2003.
8. Y. Rui, T. Huang, and S. Mehrotra, “Content-Based Image Retrieval with Relevance

   Feedback in MARS,” Proc. IEEE Int’l Conf. Image Processing, pp. 815-818, Oct. 1997.
9. K. Porkaew, K. Chakrabarti, and S. Mehrotra, “Query Refinement for Multimedia Similarity Retrieval in MARS,” Proc. ACM Int’l Multimedia Conf. (ACMMM), pp. 235-238, 1999.
10. J.R. Smith and S.F. Chang, “Visual SEEK: A Fully Automated Content-Based Image Query System,” Proc. ACM Multimedia Conf., Nov. 1996.
11. Peng-Yeng Yin, Bir Bhaneu,”Integrating Relevance Feedback Techniques for Image Retrieval Using Reinforcement Learning” IEEE Trans on Pattern analysis and Machine Intellgence, vol. 27. No 10, October 2005.